Masters Theses 1911 - February 2014

1987

The effect of background information on object identification.

Susan J. Boyce
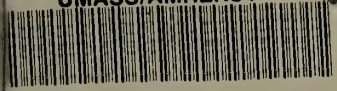*University of Massachusetts Amherst*

THE EFFECT OF BACKGROUND INFORMATION

ON OBJECT IDENTIFICATION

A Master's Thesis Presented

By

SUSAN J. BOYCE

Submitted to the Graduate School of the

University of Massachusetts in partial fulfillment

of the requirement for the degree of

MASTER OF SCIENCE

SEPTEMBER, 1987

Psychology

THE EFFECT OF BACKGROUND INFORMATION
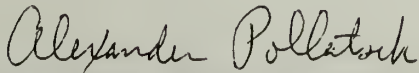
ON OBJECT IDENTIFICATION

A Master's Thesis Presented

By

SUSAN J. BOYCE
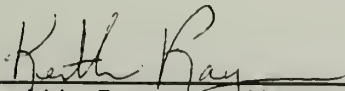
Approved as to style and content by:

Alexander Pollatsek

Alexander Pollatsek, Chairperson of Committee

Rachel Clifton

Rachel Clifton, Member

Keith Rayner

Keith Rayner, Member

Arnold Well

Arnold Well, Member

Seymour Berger

Seymour Berger, Department Chair
Department of Psychology

ii

ABSTRACT

The Role of Background on Object Identification

(September 1987)

Susan J. Boyce B.S., Ursinus College
M.S., University of Massachusetts

Directed by: Professor Alexander Pollatsek

This thesis examines the role that scene backgrounds play in object identification. Previous research has indicated that objects located in a coherent scene are easier to identify. This research employed the brief presentation method used in previous research on scene perception. Experiment 1 indicates that objects are more difficult to identify when they are located in an "episodically" inconsistent background. Experiment 2 demonstrates that the degree to which non-cued (cohort) objects are consistent with the target object has no effect on this object identification task. Experiment 3 shows that consistent episodic background information facilitates object identification and inconsistent episodic background information does not interfere. The results of these studies indicate that models of scene perception will have to be modified.

TABLE OF CONTENTS

LIST OF TABLES

C H A P T E R   1

INTRODUCTION

Experiments on the perception of pictures have suggested that the identification of an object is aided if the object is located in a "coherent scene". This work has generally focussed on degrading a coherent scene in some way in order to observe the decrement in performance. This has been done by altering the contents of the scene and by presenting the scene for brief durations to degrade the amount of information obtainable from the visual display. However, a question that has not been entirely answered is which aspects of scene coherence affect object identification.

A coherent scene consists of objects that can co-occur in the real world (objects meeting this criterion will be referred to as "episodically related"). Further, these objects must be interacting with one another in a manner that is consistent with what we know about the objects: most objects must obey physical laws and constraints due to the function of the object (e.g. a chair must obey the law of gravity and be oriented in such a way that it can be sat upon). However, even if a group of objects meets these requirements they do not necessarily constitute a scene. Well formed scenes also have backgrounds which provide information about depth and the spatial relations of the

objects. Further, the background could convey global semantic information that helps to determine the setting of the scene.

Overview

In the first section of this review I will focus on the role of scene coherence on object identification. Most of the work in this area has employed the brief display paradigm in order to degrade the quality of the scene information. Generally, the results from these studies are interpreted as evidence for rapid access of "scene schemas" that facilitate object identification.

The second section summarizes some of the literature on recognition memory for objects in coherent scenes. The theory of scene perception that emerges from this literature is rather vague, but it resembles the schema activation argument outlined by Biederman and his colleagues in the object identification section.

A brief summary of some of the research employing eye movement monitoring techniques will be reviewed in the third section. Some of the results from these studies are conflicting, but interestingly enough, the results are all interpreted as evidence for rapid schema activation facilitating object identification.

The final section of the introduction is an attempt to define the role of backgrounds in scene coherence. Scene

backgrounds have been employed in almost all research on picture processing, but little attention has been paid to the role these backgrounds may be playing. In this section I will outline some possible roles of background information.

Object Identification and Scene Coherence

Early work on object identification in coherent scenes was conducted by Biederman and his colleagues. Biederman's early work in scene perception employed photographs of scenes that were cut into sections. The sections were jumbled (resulting in incoherent scenes) or left in their appropriate positions. These coherent and jumbled scenes were presented to subjects for brief exposure durations (100-150 ms). In two such studies (Biederman, Rabinowitz, Glass and Stacey 1974; Biederman 1972) subjects were asked to identify objects in jumbled and non-jumbled scenes that were presented briefly. They found that jumbling reduced the accuracy of identifying objects, even though the objects remained intact and in their appropriate positions. However, Biederman found that the effect of jumbling was reduced if subjects were pre-cued for location of the target object and were shown the pictures of the objects before stimulus onset. Biederman et al. concluded from this line of research that when a scene was briefly presented,

getting an "overall characterization" of the scene helped to identify an object if 1) you did not know what you were looking for and 2) you did not know where to look.

Biederman's jumbled picture studies were designed to determine the degree to which destruction of the semantic coherence of the scene interfered with object identification. However, the jumbling technique disturbs more than just the overall coherence of the scene. Intact scenes were photographs of objects in their backgrounds; therefore much of the depth information was conveyed in the gradual shading differences within a background. When the picture was divided into six equal sections and put together in the jumbled configuration, non-target objects were sometimes divided between two sections. Further, jumbling introduced sharp contours that replaced gradual shading differences in background information in the non-jumbled scenes. Consequently, jumbling scenes in the manner of Biederman's early experiments does more than simply disturb the semantic coherence of the scene. The procedure results in the addition of fragments of objects and uninterpretable contours, thus not just destroying information but adding new information that was possibly disruptive.

In order to rule out the possibility that his effects in the jumbling studies were due to uninterpretable background information Biederman (1981) conducted a study aimed at determining how background information interacts

with object identification. He prepared stimuli consisting of semantically unrelated objects positioned in a depth background, a grid background, and a blank background. Subjects were presented with the sequence of events: 1) the target name, 2) the picture slide for 200 ms, and 3) a location cue. Their task was to indicate whether or not there was agreement between the target name and object at the cued location. Surprisingly, error rates were highest in the depth background condition and about the same in the no background and grid background conditions.

Biederman concluded from this result that depth information has an effect on object identification only if it aids in forming a coherent semantic representation (whatever that means). He argued that with these stimuli, no such representation could be formed because the objects were semantically unrelated, and thus that capacity used trying to construct this coherent representation took away from one's ability to identify the objects.

Another experiment (Biederman, Glass & Stacey, 1973) conducted with jumbled vs. non-jumbled pictures employed a visual search task instead of the previously mentioned rapid presentation method. The manipulations of interest were whether the object sought was or was not present in the scene and whether the object was likely or unlikely to occur in that scene. Subjects were instructed to view the picture until they found the target object or determined that the

target object was not present. High error rates were obtained in the conditions where the object was present and unlikely and where the object was likely but not present. Very low error rates were obtained in the condition where the object was unlikely to occur and was not present. Reaction times were faster overall for coherent scenes. Subjects were fastest at responding "no" when the target was unlikely to occur, next fastest for "yes" responses, and slowest for "no" when the target was likely to occur in the scene but was not present.

Biederman claimed that subjects very quickly accessed a global meaning of the scene since they made more errors when the object was present and not likely, and when the object was likely but not present. This global scene meaning, or schema, aided object identification for consistent items. Since subjects were faster at finding and identifying objects in the coherent scenes Biederman claimed that jumbling was responsible for delaying this schema activation.

Other explanations of these data exist that do not require positing schemas to aid in object identification. There is no doubt that as we move our eyes around a scene we are building a coherent representation of that scene. However, the relationship of this schema activation process to that of activating individual objects is not clear. First, it is not clear whether schema activation can occur

rapidly enough to aid the identification of individual objects. Second, even if schema activation preceeds identification of objects, it is an open question whether it influences the process of object identification. Biederman et al. data does not directly answer these questions. Subjects could have identified one or two objects in the scene and decided "no" if the target object was unlikely to be pictured with the identified objects. If it was plausible that the target object could coexist in a scene with the identified objects then the subjects kept searching. If the target object is present, they will find it before they have examined all objects in the scene (on average) and will be able to respond "yes". If the target is not present, the subjects must do an exhaustive search of all objects in order to respond "no". This strategy would account for Biederman's findings and does not require schema activation. Subjects could have just adopted a conscious strategy of deciding the likelihood of the target object appearing in the same scene as the one or two objects they initially identified.

This early research led to a flurry of experiments conducted by Biederman and his colleagues to determine more specifically when schema activation might occur and the mechanisms of this activation. Biederman, Mezzanotte, and Rabinowitz (1982) attempted to determine whether coherence and thus schema activation was due to the relationships

between the objects in a scene. They identified five "scene relations" that make up a coherent scene.

1. Support - Most objects rest on surfaces.

2. Interposition - Opaque objects occlude their background.

3. Probability - Degree to which an object is likely to occur in a given scene.

4. Position - Degree to which an object is likely to occur in a given scene at a particular spatial location.

5. Size - Degree to which an object is of an appropriate size given the size of the objects surrounding it.

Furthermore, Biederman et al. make a distinction between syntactic relations (1 and 2 above) and semantic relations (3,4, and 5 above). This distinction is based upon the claim that detection of a syntactic violation does not require identification of the object in question, and that detection of a semantic violation requires accessing object identification information. Biederman argues that a bottom up model of scene recognition (e.g. Marr, 1982) predicts that syntactic violations would be accessed by the perceptual system sooner than semantic violations, because aspects of the object (i.e spatial location) would be accessed before object identification. Thus, if lack of support was detected this would interfere at an earlier

processing stage than if a semantic relation was violated. The earlier the interference occurs in processing, the more disrupted performance should be on an object identification task.

Biederman et al. (1982) presented subjects with brief displays (150 ms) of scenes with objects undergoing various relation violations. Subjects were given the name of the target object prior to each trial and then were given a location cue and mask after the presentation of the scene. The subjects' task was to indicate if there was agreement between the object in the cued location and the name presented before the display.

Objects undergoing violations incurred a higher miss rate and a slightly higher false alarm rate than objects that were presented in their base positions. Thus, Biederman argued that objects undergoing violation are harder to perceive. However, this effect was no less for objects undergoing semantic violations than for objects undergoing violations of either support or interposition indicating that semantic violations are at least as disruptive as syntactic violations. Objects undergoing multiple violations were identified less accurately than objects that were violating only one relation.

A second experiment was conducted that involved the detection of the violations themselves. Subjects were pre-cued for the location of the potential violations and after

a brief presentation of the scene were to respond "yes" or "no" as to whether a violation was present at that location. The results were congruent with those from Experiment 1: the semantic violations were detected at least as accurately as the syntactic violations. Furthermore, multiple violations were easier to detect than single violations.

From these results Biederman concludes that objects undergoing violation are harder to see than objects not undergoing violation. Further, he claims that, since violation of a non-cued object did not interfere with object recognition for a cued non-violated object, that the elicitation of schemas for these scenes was not disrupted by the presence of a violation somewhere in the scene. In fact, he states that violation costs were the result of schema interference on object identification. As a result of this research Biederman and colleagues reject a purely bottom-up model of scene processing.

The argument against the bottom-up model relies heavily on the syntactic-semantic distinction. The result that semantic violations were as disruptive as syntactic violations seems to indicate that the bottom-up model is incorrect. However, the syntactic-semantic distinction may not be valid. In order to realize that an object is undergoing a violation of support, one has to obtain enough information about that object to determine whether it requires support. Biederman (1982) has since conceded that

object identification would be necessary to determine whether an object was undergoing the support violation. However, he counters by saying that the detection of the lack of support for an object does not require that we know what that object is. He argues that lower level assignment of surfaces during an object parsing stage would provide this kind of information. This may be true, but it does little to further the case of the syntactic-semantic distinction he is trying to defend. Identifying that an object is not supported is not the same as identifying that an object is violating the relation of support. In order to determine whether the lack of support is a violation or not still requires object identification. This same argument holds true for interposition relation as well.

The work summarized here along with other research led Biederman to believe that schema activation occurs very early in the course of picture viewing. Further, Biederman claimed that the route to schema activation is not objects-then-schema as has been proposed by Friedman (1979) and others. The route that he advocated requires the use of the relations or interactions between the objects of the scene and he asserts that this information is available before the objects are fully identified.

The research conducted by Biederman and his colleagues does indicate that subjects are able to gain some information from very brief exposures to pictures. His

evidence for accessing relations between objects before the objects themselves have been identified seems weak. However, these experiments all clearly indicate that when scene coherence is destroyed subjects are less likely to recognize a target object in a scene. Experiment 2 in the present set of experiments will test Biederman's route to schema activation.

Object Recognition and Scene Coherence

Antes and various colleagues have also attempted trying to sort out the kinds of information a viewer can get from brief displays of pictures. Antes (1977) presented subjects with 100 ms displays of scenes then tested subjects for recognition of sections of the scene. The sections were created by dividing the picture into equal parts (much as Biederman had done) and the sections were rated as being of high, medium, or low "informativeness". He found that accuracy on the section recognition task depended on the rated informativeness of the section (subjects were more accurate for highly informative sections) and on the eccentricity of the section (subjects were more accurate if the test section had been located near the center of the picture). A second experiment employing the same methodology tested subjects' ability to recognize the location of a target section. Subjects were presented with

the scene and then a target section of that scene. Their task was to indicate where in the scene that section had appeared. Accuracy on this task was lower overall than on the section recognition task but performance was still affected by the informativeness of the section and by the eccentricity of the target section. Since subjects were more accurate at recognizing the section than localizing the section Antes proposed that object identification and object localization are mediated by two different processes and that the process responsible for localization is slower. The evidence for this model seems pretty weak.

In an experiment geared toward identifying global versus local processing of scenes, Antes, Penland, and Metzger (1981) presented subjects with brief displays of scenes followed by an object recognition task. The manipulations of interest were scene context (high or low), usualness of target object (usual or unusual with respect to context of scene), and consistency of distractor information (consistent or inconsistent with respect to context of scene). "High context" scenes were well formed scenes with coherent backgrounds and many objects, while "low context" scenes were created by deleting the background and some of the objects from the high context scenes. Subjects saw a 100 ms exposure of the scene followed by an array of four objects (three distractor objects and one that had appeared in the scene). Subjects were instructed to choose the

object that they had seen in the display.

For high context scenes subjects were most accurate if the target object was consistent with the context and the distractor objects were inconsistent with the context. Subjects were less accurate if the distractor objects were also consistent with scene context and performance was poor when the target object was unusual with respect to the context of the scene. Thus, probability of correct recognition of an object was influenced by both usualness of the target and usualness of the distractors. In contrast, performance of subjects presented with low context "scenes" were not affected by the usualness of the target object or consistency of distractor objects. Antes et al. ran a third group of subjects that were not presented with the scene at all but instead read a one sentence discriptor of the theme of the scene. These subjects then chose an object from the array of four objects that was likely to occur in a scene with the particular label. The pattern of data obtained from this "thematic information group" is very similar to the data collected from the high context group.

From this study Antes et al. conclude that global scene information is available after 100 ms of viewing a scene. Because the results from the thematic information group parallel those from the high context group, Antes advocates a "scene emergent features" route to schema activation. According to this theory, the first fixation on the scene

provides global information about the setting, thus allowing the viewer to call upon knowledge about which objects to expect in the scene. His evidence against the objects-then-schema route is the similarity between the thematic information group and the high context group. If schema activation occurs as the result of identifying one or two objects in the scene, then the high context unusual target object group should have accessed the wrong schema on trials where the unusual object was identified first. This should create a discrepancy between the high context group and the thematic information group and this descrepancy was not evident in the data.

Unfortunately, Antes is not able to define what these "scene emergent features" might be. Biederman defines them in terms of relations between objects. But if one is to believe the results of Antes (1977), Information about an object's location is accessed only after object identification has occurred. This leaves a somewhat muddled picture about how context helps in picture viewing.

Eye Movement Research and Scene Perception

Research on the role of context on object identification has not been limited to paradigms that employ brief presentation of scenes. Many researchers have recorded eye movements during scene viewing to determine the

role of scene context on a subject's fixation duration and pattern of fixations. Loftus and Mackworth (1978) were interested in determining where people will look in a scene given the opportunity to view the display for 4 seconds. Some of the scenes they presented contained objects that were very unlikely to occur given the scene context (i.e. octopus in a farm background, tractor in an underwater background). They found that low probability objects were fixated earlier, were fixated more often, and were fixated for longer durations. From these data Loftus and Mackworth claim that subjects readily obtain the "gist" of the scene (within the first fixation) and partially identify objects in the periphery. This partial object identification then leads subjects to compute conditional probabilities that these objects are likely to occur given the gist. Subsequent fixations are then presumably guided to those objects whose conditional probabilities are lowest.

The examples of the stimuli used in this experiment, if representative of the stimuli set as a whole, provide an alternative explanation of these data. The octopus in the farm background is not only semantically or episodically different from the rest of the scene but is also different in terms of low level perceptual features. The farm background consists predominantly of straight lines and right angles, whereas the octopus is defined by irregular contours and wavy lines. Subjects could have fixated the

octopus earlier and for longer durations because of this contour difference and not because they were being influenced by the semantics of the object.

If the stimuli presented in the article were not representative of their stimuli in general, and their results really do indicate that subjects will look to objects that are not probable, then the route to obtaining the gist of the scene must be rapid processing of the background information. This would argue that a scene that either had no background, or a noninformative background would be processed in a different manner than that outlined by Loftus and Mackworth.

Antes and Penland (1981) conducted a similar study; they recorded eye movements during viewing of scenes that contained high probability and low probability objects. An additional manipulation was included in this study: the degree of background context. High context and low context scenes were defined as in the Antes, Penland and Metzger (1981) study described above. They found that first fixations on expected objects were shorter when they occurred in high context than when they occurred in low context. First fixation durations on unexpected objects were the same in both the high context scenes and the low context scenes. Data collected on the average saccade length indicated superior peripheral identification of objects that were consistent with the context than of objects that were

inconsistent with the context. Concerning the probability of fixating a likely object versus an unlikely object, little can be said from this study, because probability of an object occurring in the scene was confounded with location in the scene so that more unexpected objects occurred in the center of the scene where subjects began viewing. An attempt was made to sort out this confounding and Antes and Penland concluded that objects consistent with the context were more likely to be fixated early in viewing while during later viewing the pattern switches so that unexpected objects are more likely to be fixated.

Both Loftus and Mackworth (1978) and Antes and Penland (1981) argue that the theme or gist of the scene can be obtained very rapidly which results in activating schemas for these scenes. It is interesting to point out that this concept of schema activation does not constrain the predictions that can be made concerning the likelihood of a given object being fixated. Loftus and Mackworth claim that this schema activation results in fewer fixations on likely objects while Antes and Penland argue that schema activation results in more fixations on expected objects. Perhaps a better understanding of how schemas are activated may make clearer what schemas are and may help constrain the theory.

Various other studies have been conducted using eye movements to investigate scene perception (Mackworth and Mirandi, 1967; Nelson and Loftus, 1980; and Saida and ikeda,

1979). Results of these studies are in many ways conflicting in that the schema argument can be used to predict many outcomes. Various other people have studied memory for pictures (Friedman, 1979; Parker, 1978). This literature will not be reviewed at this time since the issues that are dealt with in recognition memory for pictures may be quite different than those concerning identification of scenes.

A Definition of "Coherent Scene" Inferred from Stimuli and Manipulations in Previous Research

The research reviewed above, when taken together as a whole, appears to share a common definition of "coherent scene". The coherent scenes that have been used in these experiments have common elements. First, coherent scenes always consist of objects that are episodically related; meaning that the objects could co-occur in a real world scene. Second, these objects must maintain certain physical relationships with one another as Biederman et al. (1982) have indicated. Objects must obey physical laws (i.e. gravity, opaque objects occlude their backgrounds, etc.). Furthermore, objects must be placed in the scene in such a way as to establish a consistent viewpoint (i.e. objects further away look smaller). Third, well formed scenes have a "background". This third criterion for well formed scenes

will be the focus for the rest of the paper.

The Role of Backgrounds In Coherent Scenes

The well formed scenes employed in the experiments reviewed above have all included some type of background information, yet surprisingly little has been said about the role this information plays in achieving the meaning of a scene.

Biederman (1981) touched on the role of backgrounds in his investigation of depth gradients and object identification. Depth gradients actually interfered with one's ability to identify objects. However, the scenes that he used were not coherent. The objects violated episodic relatedness and the backgrounds were merely converging lines constructed to establish a viewpoint for the subjects. Biederman concluded that background information (or more precisely, depth information) should help only if the objects are episodically related, but this fact has not been established.

Antes, Penland, and Metzger (1981) also tangentially dealt with the role of backgrounds by nature of the way they defined high and low context pictures. The high context scenes used in this study contained many objects (some more defining of the scene than others) as well as fairly complex

backgrounds. The low context scenes were created by deleting the background and some of the objects from the high context scenes. Antes et al. found object recognition performance was much better in the high context scenes than in low context. Unfortunately, because some objects had been removed as well as the background to create low context scenes, no conclusion can be drawn concerning the role that the background itself plays in providing context (this was also true of Antes and Penland, 1980, reviewed above). Furthermore, because Antes did not report how he chose the objects to be deleted, it is possible that some context defining objects (objects that can only fit in that scene) were deleted, which would create a decrement in performance.

Potential Roles of Background Information on Object Identification

Background information may provide context by helping to establish a unique viewpoint for the scene and by providing depth cues. By unique viewpoint I mean establishing whether it is a top view, or a view from the left corner of the room etc. This is much the way Biederman expected his depth gradients to function. It is possible that some of the relations between objects that Biederman has identified are dependent upon background functioning in this manner. For example, violations of relations of size

and support should be very difficult (maybe impossible) to detect if background depth information is missing. It is not clear how one can, without depth information, determine if an object is in an appropriate place and if it is the correct size. It may be possible to construct depth information from the objects if none is given by the background, but this process would take much more time.

Another way in which backgrounds may function is by providing "episodic cues": the background provides some information as to the theme or meaning of the scene. For example, a Farm Background may contain a barn in the distance, fences around fields, and perhaps rolling hills in the distance. In this regard, backgrounds may function as large objects with recognizable features. Perhaps what subjects obtain from brief presentations of scenes is background information that then allows them to infer the objects that are likely to occur in that background.

The experiments reviewed above indicate that scene context does alter our ability to correctly identify and accurately remember objects. When scenes are coherent identification of objects appears to be facilitated and we are more likely to remember objects that were consistent with the context. However, it is not clear from the above studies the degree to which this apparent facilitation from scene context is due to critical objects in fovea, or critical aspects of the background. The question left

unanswered is the degree to which benefits from scene context are due to the episodic consistency of the backgrounds.

The experiments reported here are designed to examine the role of background information on the identification of objects. Experiment 1 establishes whether depth gradient information alone is sufficient to facilitate object identification or whether episodic, setting information is obtained from brief displays of scenes.

All three experiments employed the brief presentation method used by Biederman and others. Conceptually, this task is trying to mimic what might occur during the first 150 ms of scene processing, or roughly the first fixation on a scene. There are a number of reasons why brief exposures may not mimic first fixations. First, when subjects know they are only going to have 150 ms of viewing time they may allocate spatial attention differently. Perhaps an attempt is made to "widen" their spotlight of attention to take in as much of the display as possible and, to the extent that they are successful in this, there may be a decrement in processing the foveal information. Another possible problem with this task is that conclusions about the time course of processing are not straightforward. The visual display is terminated 150 ms into a trial and subjects' responses (yes or no in the object identification task) generally are produced about 1000 ms after the display is terminated.

This 1000 ms response time is not only reflecting the time to engage in a motor movement (for a button press or vocal response) but also reflects decision time, or "post-perceptual" processing.

The decision to use the brief presentation method was made knowing that the procedure was flawed in some respects. This method has been commonly used in picture perception research so that it makes comparisons between the current studies and previous research relatively easy. Also, to some extent, all methodologies have flaws. Even eye movement research (which is by far the most on-line measure available) is not without associated problems. It could be debated that no measure from the eye movement record, such as first fixation duration or gaze duration is an uncontaminated measure of the time to identify an object. A final reason for choosing the rapid presentation method was one based on pragmatics. This method was relatively easy to implement and the equipment was available.

Overview of Experiment 1

The manipulation of interest in Experiment 1 was whether the object to be identified was presented in an episodically consistent background, an episodically inconsistent background or with no background at all.

Background consistent scenes were coherent scenes with the background conveying information about the setting of

the scene. The objects located in the foreground were objects that are likely to occur in that setting. The objects were positioned in the background in such a way that does not violate any of Biederman's relations.

Background Inconsistent scenes consisted of well defined backgrounds that convey information about the setting of the scene. The objects located in the foreground were objects that are episodically related to one another, but could not occur in the background. The objects were positioned within the background so as not to violate relations of size or support. Two no background control conditions were included in order to establish a baseline for the object identification task. These no background condition "scenes" were created by deleting the background from the scenes in each of the two above conditions.

Biederman's model predicts that objects in the consistent background scenes would be perceived more readily than objects in the inconsistent background scenes. Further, Biederman's model states that the route to schema activation (and therefore facilitation in object identification) is through accessing relations between objects. Thus performance should be worse in the no background conditions than in the background consistent conditions, as no background scenes make it very difficult (maybe impossible) to identify the relations between the objects.

The bottom-up model of object identification, proposed by Henderson, Pollatsek and Rayner (1987) posits that scene context effects may be due to passive spreading of activation between "object" nodes in a network. Taking this position to the extreme, one would predict no difference between any of the conditions in experiment 1 since the target object is always presented with related objects. However, if backgrounds function as "large objects" and are represented in the network as nodes the same as other objects in the scene, then one would predict that performance on background consistent scenes would be better than performance on background inconsistent scenes.

C H A P T E R   2

EXPERIMENT 1

Introduction

Experiment 1 was designed to investigate the role backgrounds play in object identification. Scenes were constructed so that objects appeared either in episodically consistent backgrounds, inconsistent backgrounds or no backgrounds at all. The first purpose of this experiment was to determine if the degree of episodic relatedness had an effect on object identification, and second, whether consistent backgrounds facilitated object identification or inconsistent backgrounds interfered with object identification, or both.

Method

Subjects

Sixteen University of Massachusetts undergraduates participated in this experiment for extra credit in psychology courses. An experimental session lasted approximately 55 minutes.

Scenes

The 64 scenes were constructed from an original set of

16 coherent scenes. These 16 scenes were line drawings of common rooms of a house (e.g. Bedroom scene), common public places (e.g. Diner/lunch counter), and common outdoor scenes (e.g. Street scene) (see Appendix for a list and description of scenes). Each of the original 16 scenes consisted of a background and five objects and they were organized in 8 scene pairs with the constraint that objects in both scenes in the pair be roughly equivalent in real-world size. The 16 Background Inconsistent scenes were created in the following manner: The objects in one scene were switched for the objects in the paired scene. The placement of objects in the inconsistent background required that the objects be rearranged in the two dimensional frame in order not to violate support, and in no case was the object's retinal size altered. The two no background control scenes were created by eliminating the background from the 16 original scenes and the 16 background inconsistent scenes. The scenes subtended 16 degrees in width and 14 degrees in height and the objects averaged approximately 2 degrees in width and 2 degrees in height and were located 5 degrees on average from the fixation point.

Design

The 64 scenes were created from 16 backgrounds and 16 object sets. Each subject saw each object set four times (one time for each background condition). However, a

different object in the set was cued each time. For example, the kitchen sink object set consisted of a coffee maker, wine glass, toaster, eggbeater, and a bottle of dish washing liquid. One of the five objects, the dish washing liquid in this case, was never cued. Selection of the object not to be cued was based upon the difficulty of labeling it and the degree to which the object was distinguishable as belonging in that scene. One subject was cued for the coffee maker in the background consistent condition, the toaster in the background inconsistent condition, etc. This was counterbalanced across subjects within a given scene and within a subject across the different scenes. The correct answer (yes or no) was also counterbalanced so that for every subject half the correct answers were yes and half were no. Therefore, across a set of 8 subjects, all four objects would be tested in each background condition, half of the time when the object name was presented and half the time when when the name of an object not in the scene was presented. Two names of objects not present were needed for each scene for the no trials. These object names were chosen so that they were likely to fit in with the background (list of these names can be seen in the Appendix, labeled "distractor items"). Trials were presented in a random order, with the constraint that no two consecutive trials contained the same background or the same object set.

Apparatus

The scenes were displayed on a Megatek Whizzard Vector plotting CRT scope with a P-31 phosphor interfaced with a VAX 11-730 computer. The computer controlled the experiment and recorded the subjects' responses and response times. The scenes were entered into the computer by a Summagraphics Bit-Pad.

Procedure

Subjects first read the name of the target object that appeared in the center of the screen in front of them. The name remained on the screen for 3 seconds and was immediately followed by a fixation cross in the center of the screen. The fixation cross remained on the screen until the subject initiated the trial by pressing a response key. The scene appeared on the screen for 150 ms and was followed immediately by a mask (consisting of random line segments and angles) with a "cue" embedded in it. The cue was a filled circle approximately 1/2 in diameter. The location of the cue varied widely, but it was always where an object had been present in the scene. The subject was instructed to respond yes if the name presented at the beginning of the trial was the name of the object present at the cued location in the scene. Subjects were to respond no if target name was not the object at the given location in the

scene. Subjects' <u>yes</u> and <u>no</u> responses were made by pressing a response key with either the middle or index finger of their right hand, respectively. After the subject had responded, the mask was removed from the screen. The next trial began with the target name presentation 8 seconds after the previous response had been made. The 128 experimental trials were divided into 4 sets and subjects were given a 3 minute break between each set. The first two sets composed the first time through the 64 scenes (and will be referred as Block 1) and the second two sets were the 64 scenes shown for a second time (Block 2). 32 practice trials (using different scenes than those in the experimental trials) were given to each subject to ensure that they understood the task.

Results

Dependent Measures

The goal of Experiment 1 was to determine what aspects of background information influence accuracy on the object identification task. However, there is no single best measure of accuracy. One method frequently used is to look at the probability of hits (<u>yes</u> when object was present) and the probability of false alarms (<u>yes</u> when object was not present). If the probability of a false alarm decreases while the probability of a hit increases then assessing

accuracy is pretty straightfoward. However, if an increase in the probability of a hit is associated with an increase in the probability of a false alarm, then assessing accuracy or "sensitivity" is more of a problem. In all the experiments reported here false alarm rate increased with the hit rate. In some sense, "percent correct", averaging over yes and no trials, corrects for the response bias problem. Other more theoretically based measures of sensitivity have been proposed, however. d' and A' measures, two such measures designed to control for response bias, were also used in the current experiments. A' was used because it has the capability of dealing with probabilities of 0 and 1 in the false alarm and hit rates, repectively (see Grier, 1971, for details). In general, the pattern of results in all three experiments was the same across all measures. Therefore, I will concentrate on Percent Correct because it is easiest to interpret.

In all of the analyses to follow there was a significant main effect of block (i.e. subjects made fewer errors in the second block). However, since the block factor never interacted with the other factors, data from both blocks have been combined and block has been dropped as a factor.

Although Response Times were recorded the results will not be reported. The times were relatively long and extremely variable, so they have been dropped from

consideration.

Accuracy

The primary goal of Experiment 1 was to assess the degree to which backgrounds influenced accuracy on the object identification task. Theoretically, this influence could be measured in the difference between the <u>Background Consistent Condition</u> and the <u>Background Inconsistent Condition</u>. However, since the objects were in different orientations in these two conditions we cannot assess the difference simply by looking at the main effect of Consistency. The interaction between Background Presence and Consistency is the appropriate measure of background effect, since each background condition has a no background condition matched for location of the objects.

Table 1

<u>Mean</u> <u>Percent</u> <u>Correct</u> <u>and</u> <u>Mean</u> <u>D'</u> <u>on</u> <u>Object</u> <u>Identification</u> <u>Task</u> <u>in</u> <u>Experiment</u> <u>1</u>

	Background Consistent		Background Inconsistent	
Background Present	66.3%	(.901)	58.8%	(.563)
Background Absent	67.2%	(.901)	67.8%	(1.04)

<u>Note</u> - D' in parentheses.

The overall percent correct was 64.04%. As can be seen in Table 1, the difference between the Consistent Background condition and its No Background control was 0.9%, while the difference between the Background Inconsistent condition and its No Background control was 9.0%. This Consistency x Background Presence interaction was significant, $F(1,15) = 5.95$, $p = .028$ in a two factor repeated measures Analysis of Variance. The Consistency x Background Presence interaction was significant for the d' measure as well as the A' measure, $F(1,15) = 6.29$, $p = .024$, and $F(1,15) = 5.11$, $p = .039$, respectively.

A secondary question is whether the relatively good performance in the Background Consistent condition is the result of facilitation, or the relatively poor performance in the Background Inconsistent condition is the result of interference. In order to address this question simple effects t-tests were conducted on the difference between each background condition and its matching no background condition. The difference between the Background Consistent condition and its No Background control was not significant, $t(15) = .414$, $p > .05$, while the difference between the Background Inconsistent and its No Background control was significant, $t(15) = 3.87$, $p < .01$.

Finally, lets briefly look at the False Alarm and Miss data to demonstrate the trade - off between Hits and False Alarms.

Error Data - Misses and False Alarms

The percentage of misses (subject responds no when target object was cued) and the percentage of false alarms (subject responds yes when target object was not cued) can be seen in Table 2. Overall, subjects missed on 40.58% of the yes trials and false alarmed on 29.6% of the no trials. Both the miss and the false alarm data were submitted to a two way analysis of variance. The miss results will be discussed first.

Table 2

Mean Percent of Misses and False Alarms for Experiment 1

	Background Consistent		Background Inconsistent	
Background Present	36.8%	(30.9%)	59.4%	(22.7%)
Background Absent	36.8%	(29.7%)	29.3%	(35.2%)

Note - False alarms in parentheses.

The main effect of Consistency approached significance with $F(1,15) = 4.21$, $p = .058$. The Presence/Absence of Background factor was significant, $F(1,15) = 16.79$, $p = .001$

and the Consistency X Background Presence Interaction was highly significant, $F(1,15) = 25.50$, $p = .0001$.

A two-factor within subjects ANOVA was performed also on the false alarm data. The background Presence factor was marginally significant, $F(1,15) = 3.03$, $p = .103$ and there was a significant Interaction between Consistency and Background Presence, $F(1,15) = 6.54$, $p = .022$.

As can be seen in Table 2, the drastic increase in misses in the Background Inconsistent condition is accompanied by a less dramatic, but significant, decrease in the false alarms in the Background Inconsistent condition. This trade-off suggests that subjects' criterion for responding 'no' had shifted so that overall they responded 'no' more often in the Background Inconsistent condition, regardless of the presence or absence of the target object.

Discussion

Analyses of all the accuracy measures Indicated that subjects performed better on the object Identification task when the background was either consistent with the objects or the background was not present at all. First I will discuss the difference between the background consistent condition and the background inconsistent condition. Second I will address the results of the no background conditions.

As outlined in the Introduction, I proposed that

backgrounds may serve two purposes: to establish depth and size cues, and to provide episodic information (theme or meaning of scene). The background inconsistent scenes in this experiment were constructed in such a manner as to preserve the first function of backgrounds while violating the second. It is clear from the pattern of both the percent correct data and the d'/A' data that the episodic information provided by the background influenced the subjects' responses on the object identification task. From this we can conclude that backgrounds do more than just provide the appropriate depth and size relationships; they are also important in establishing the meaning of the scene. Furthermore, the data indicate that the process of extracting the meaning from the background can be done very quickly.

It is important to note that the influence of the backgrounds in this study was not a result of some diagnostic features in the fovea. The portion of the scenes that fell on the subject's fovea contained no objects and contained no diagnostic background features (e.g. a faucet in kitchen sink scene).

The benefit in object identification provided by the consistent background seems to conflict with a strong interpretation of the priming model outlined by Henderson, Pollatsek and Rayner (1987). The intralevel priming hypothesis posits that context effects in scene perception

can be accounted for by an object to object priming mechanism. Since the target object in the current experiment always appeared in scenes with other related objects, the degree to which the target object was primed should have been constant across all conditions. However, if we relax our definition of what can be an object in the network and conceive of backgrounds as functioning as large objects then the results obtained in this experiment are not incongruent.

The finding that subjects were able to identify objects in scenes that had no background as well as they were able to identify objects in scenes that had a consistent background may pose a problem for Biederman's object relations route to schema activation. Biederman et al. (1982) claim that a schema for the scene is activated on the basis of accessing information about the relationship between the objects in the scene. As I have pointed out in the Introduction, information about the relationships between objects would not be available (or at least only partially available) in the scenes that had no background. In the no background scenes employed in this experiment, all the objects lacked support and violated most of the other Biederman relations, which are defined in terms of the object and the background. These violations should have made performance much worse in the no background condition as compared to the background consistent condition if the

relation information was accessing a schema and object identification was facilitated by this schema. The pattern of results indicate that this did not happen.

It is possible that subjects' equivalent performance in the background consistent condition and the no background conditions was produced by different, off-setting factors. Good performance in the background consistent condition could have been due to facilitation from the background. On the other hand, the good performance in the no background condition could have been the result of these scenes being "perceptually easier" since the no background scenes were considerably less complex than the background scenes. If the first stage in scene perception is to identify where the objects are in the scene and then parse the object from the background (i.e. decide which lines belong to the background and which belong to the object), then it does not seem unreasonable that performance in the no background conditions was relatively good. When the no background scenes were presented, the initial parsing stage of processing the scene had already been done, so subjects may have had longer to identify the objects.

Experiment 1 indicated that background information can be utilized from a brief (150 ms.) display of a scene. However, many questions remain. In Experiment 1 the target object was always presented along with related objects. As a result, it cannot address the role that these non-cued

objects might have played. Furthermore, we are unable to determine precisely why performance in the no background condition was as good as performance in the consistent background condition. The next two experiments will address these issues.

C H A P T E R   3

EXPERIMENT 2

Introduction

Experiment 1 demonstrated that background information influenced performance on the object identification task, even though the subject's exposure to the scene was brief. In Experiment 1, however, all scenes consisted of a target object with related objects in one of the background conditions. Because objects were presented as a set, it is not clear what effect the non-cued objects had on identification of the cued object. Henderson, Pollatsek and Rayner (1987) have found evidence for objects priming related objects. It is possible that in Experiment 1 non-cued objects did prime identification of the cued object, but that the effect of this priming was constant across all conditions.

Experiment 2 is an attempt to sort out the effects of non-cued objects from the background effects. Possibly, both related objects and consistent background facilitate object identification. In this experiment subjects were presented with scenes that either had a consistent background, an inconsistent background or no background, as in Experiment 1. In contrast, the target object in Experiment 2 was presented either with four related objects or with four unrelated objects. In this way the

41

independent roles of the cohort object set and background can be tested.

Method

Subjects

Sixteen different University of Massachusetts undergraduates participated in this experiment for extra credit in psychology courses. An experimental session lasted approximately 55 minutes.

Scenes

The 128 scenes were composed from the same original 16 scenes as in Experiment 1. These 16 scenes were grouped into 8 pairs as outlined in Experiment 1. The objects in one scene of the pair were switched with objects in the other scene pair resulting in 16 inconsistent background scenes. In order to test the role of cohort objects, one object was selected from each scene and switched into the inconsistent background, leaving the other four objects consistent with the background. An example may help to illustrate this: The <u>Bedroom</u> <u>scene</u> and the <u>Refrigerator</u> <u>scene</u> were paired together. In the <u>Background</u> <u>Consistent</u> condition bedroom objects appeared with the bedroom background. In the <u>Background</u> <u>Inconsistent</u> condition, bedroom objects appeared with the refrigerator background (so far this is the same as Experiment 1). In the

background consistent/target not with cohorts condition the doll (bedroom object) appeared in the bedroom background with four refrigerator objects. In the background inconsistent/target not with cohorts condition the suitcase (bedroom object) appeared in the refrigerator with four refrigerator objects. Each of these four conditions had matching no background conditions (as in Experiment 1) resulting in 8 conditions (see Table 3 for summary of the scene conditions).

Table 3

Example of Scene as it Appears in the Conditions of Experiment 2.

	Background Consistent	No background Consistent
With Cohorts	Target = Doll Bedroom background with bedroom objects	Target = Doll No background with bedroom objects
Not With Cohorts	Target = Doll Bedroom background with fridge objects	Target = Doll No background with fridge objects

	Background inconsistent	No background inconsistent
With Cohorts	Target = Suitcase Fridge background with bedroom objects	Target = Suitcase No background with bedroom objects
Not With Cohorts	Target = Suitcase Fridge background with fridge objects	Target = Suitcase No background with fridge objects

In order to avoid cuing the same target object in each of the eight conditions, two target objects were selected from each object set (doll and suitcase in the example above). These objects were selected on the basis of an analysis of the items from Experiment 1. All target items selected were identified correctly 75% of the time in Experiment 1. Objects were positioned in the background in the same places as in Experiment 1. The size of the scenes, size of the objects and distance between objects and the fixation point was the same as in Experiment 1.

Design

Each subject saw each of the 128 scenes one time. However, because of the overlap between scenes each subject saw each background 4 times, each set of non-cued objects four times and each target object 4 times. On half the trials the target object name was presented and on half the trials the name of an object not in the scene was presented. The correct answer (yes or no) was counterbalanced across conditions. Trials were presented in a random order, with the restriction that no two consecutive trials had the same target object.

Apparatus and Procedure

The apparatus and procedure employed in this experiment

were the same as in Experiment 1.

Results

As in Experiment 1, a trade-off between Hits and False Alarms required that some measure of sensitivity be computed. In the current experiment, some subjects had hit rates of 1.00 and false alarm rates of 0.00 in some conditions. For this reason d' cannot be computed. However, the non-parametric measure of sensitivity, A', can deal with probabilities of 0 and 1. A' can be interpreted as the best estimate of what the percent correct would have been if a forced-choice procedure had been employed instead of a presence-absence procedure.

Accuracy

As with Experiment 1 the appropriate measure of the effect of Background Consistency is the Background Presence x Background Consistency Interaction. The overall percent correct was 65.45%.

Table 4

Mean Percent Correct in Experiment 2.

	Background Consistent	No Background Consistent
With Cohorts	66.4%	68.8%
Not With Cohorts	69.5%	73.5%

	Background Inconsistent	No Background Inconsistent
With Cohorts	57.1%	65.4%
Not With Cohorts	54.9%	68.0%

Table 4 displays the percent correct for each condition. Collapsing across Cohort for the moment, it can be seen that the effects of Experiment 1 have been replicated. The difference between the Background Consistent condition and its matching No Background condition was 3.2%, while the difference between the Background Inconsistent and its No Background condition was 10.7%. This interaction was significant, $F(1,15) = 6.47$, p = .0225. The pattern of data for the A' measure was

similar, but the Consistency x Background Presence interaction was only marginally significant, $F(1,15) = 2.97$, $p = .1054$.

Table 5

Mean A' for Experiment 2.

	Background Consistent	Background Inconsistent
Background Present	.748	.602
Background Absent	.795	.741

To determine the effect of cohort set, two results might be of importance. If it is assumed that the role cohorts play in object identification is independent of background, then a main effect of the Cohort factor should be evident. However, if cohort set affects object identification differently when the background is present or absent, or when the background is consistent or inconsistent then a Cohort x Background Presence x Consistency interaction would be present in the data. A three factor within subjects ANOVA performed on the data indicated that there was no effect of Cohort set, either as a main effect, $F < 1$, or interacting with Background Presence and

Consistency, F < 1. There was no effect of cohort evident in the A' data as well.

Error Data — Misses and False Alarms

The percentage of misses and the percentage of false alarms can be seen in Table 6. Overall, subjects missed on 35.36% of the yes trials and false alarmed on 33.6% of the no trials. Both the miss and false alarm data were submitted to a three way Analysis of Variance. The miss results will be discussed first.

The main effect of Consistency was significant $F(1,15)$ = 34.33, p = .0000, as well as the main effect of Background Presence $F(1,15)$ = 25.56, p = .0001. There was no hint in the miss data of any effect from the Cohort factor. The Consistency x Background Presence Interaction was significant, $F(1,15)$ = 28.53, p = .0001.

The pattern of the false alarm data looks slightly different than the miss data. In the false alarm data there was a main effect of Target With Cohorts, $F(1,15)$ = 8.05, p = .0125. False alarms increased when the cued object appeared with objects that were plausibly cohorts. The only other significant effect was the Consistency x Background Presence Interaction $F(1,15)$ = 8.64, p = .0102.

Table 6

Mean Percent Misses and False Alarms for Experiment 2.

	Background Consistent		No Background Consistent	
With Cohorts	24.6%	(41.2%)	25.0%	(37.3%)
Not With Cohorts	23.5%	(36.1%)	23.5%	(28.4%)
	Background Inconsistent		No Background Inconsistent	
With Cohorts	57.8%	(29.0%)	29.7%	(40.6%)
Not With Cohorts	66.6%	(22.8%)	31.1%	(33.4%)

Discussion

The percent correct data and the A prime data (to a lesser degree) replicate Experiment 1. Subjects performed equally well in the no background conditions as in the background consistent conditions. Furthermore, performance dropped when the background was inconsistent with the target object as compared to the background consistent condition where the background was consistent with the target object.

Perhaps the most surprising result from Experiment 2 is that it made no difference whether the target object was surrounded by consistent or inconsistent cohort objects. It is particularly interesting that, even in the no background conditions, there was no effect of cohort set. It seemed likely that without background to aid in obtaining the "theme" of the scene, subjects might be able to utilize information from the surrounding objects more. This was not the case. First I would like to discuss the ramifications of these results for Biederman's object relations model. Then, I will discuss the implications of these results for the object to object priming model outlined by Henderson et al.

Biederman et al. (1982) claimed that the route to schema activation was through identification of the relationships between objects. In the current experiment, the target object violated the probability relation in some conditions (the degree to which an object is likely to occur in the scene) no matter how one defines probability (with respect to the background, with respect to the other objects, or with respect to both). Even in the strongest case of violation, where the background and all non-cued objects were from the same scene and only the target object was inconsistent, only an effect of background was obtained.

This calls into question Biederman's route to schema activation. The information subjects were obtaining from

scenes where the target object was not with its cohorts should have been degraded and schema activation should have been slowed or possibly inhibited altogether. The discrepancy between the current data and those obtained by Biederman et al. is particularly surprising since care was taken to ensure that the object identification task employed in this experiment was the same as the task used by Biederman et al. I have essentially not replicated Experiment 1 of Biederman, Mezzanotte, and Rabinowitz (1982).

Experiment 1 indicated that the Henderson et al. object to object priming model would have to be modified: if the background functions as a large object priming other objects then, a priming explanation can account for the decrement in performance when the background is inconsistent. However, even this modified version of this model cannot readily account for the lack of effect of cohorts in this experiment. According to the priming mechanism, we should have seen a benefit when the target object was presented with cohorts as opposed to being presented with unrelated objects.

There are some possible explanations as to why Henderson et al's model does not fit these data. One possibility is that the task used in this experiment taps processing at a considerably later stage. Henderson et al's object priming effects were reflected in fixation durations,

while the subject had considerably more time to look at the object. Subjects in the current experiment saw the scene for only 150 ms. and never had the object in their foveal region. Probably the most important difference between Henderson's priming paradigm and the current methodology is the difference between objects being presented foveally or parafoveally. The current object identification task may have been tapping some reconstructive process that subjects engaged in after the scene was no longer present. This possibility will be discussed in more detail in the General Discussion.

Another possible explanation for the lack of priming effects in the current experiment has to do with what should prime what. Henderson et al's priming was achieved with objects that were semantically related (i.e. Doctor - Nurse) and not necessarily episodically related. This is not to say that Doctors and Nurses cannot co-occur in the same scene, because of course they can. However, the stimuli in the current experiment were always episodically related but many times were not semantically related (i.e. Doll - Suitcase). Both a doll and a suitcase can easily occur together in a bedroom scene, but they are not semantically related. Taken out of the context of the bedroom, one would predict little priming of doll from suitcase and vice versa. This too will be discussed further in the General Discussion.

In conclusion, it appears that the degree to which the background is consistent with the target object has a rather large effect, whereas the degree to which the non-cued objects are consistent with the target object has little or no effect.

C H A P T E R   4

EXPERIMENT 3

<u>Introduction</u>

At this point we know that the degree to which a background is consistent with the target object predicts performance on the object identification task. However, we do not know precisely in what manner the backgrounds are operating. If we assume that having a consistent background facilitates object identification, then there is a problem reconciling the results from the no background conditions. If facilitation from the appropriate background aids object identification, then performance in the no background control conditions should have been worse than in the background consistent condition. However, to accept the no background controls as the appropriate baseline in this object identification task one would have to argue that there is no facilitation from the consistent backgrounds, only interference from processing an inconsistent background.

There is reason to believe that the no background conditions are really not the appropriate control condition. As outlined in the discussion of Experiment 1, much is left uncontrolled in the no background condition. No background conditions not only lack the "theme" information that background conditions have, but they are also less complex.

54

Subjects may be able to identify more objects in the no background conditions because they do not have to first identify where the objects are and parse the objects from the background. A better control would be one that preserved background complexity, yet had no real meaning. The purpose of Experiment 3 was to test the consistent and inconsistent background effects against a more appropriate control condition.

Method

Subjects

Twenty-four different University of Massachusetts undergraduates and University of Massachusetts graduate students in psychology participated in this experiment. The undergraduates received extra credit in psychology courses for their participation. An experimental session lasted approximately 55 minutes.

Scenes

The same 16 original scenes were employed as the consistent background scenes. Inconsistent background scenes were created from these original 16 scenes in the same manner as Experiment 1. Instead of no background control conditions, Nonsense Backgrounds were used as a control. Nonsense backgrounds were created by distorting

the original backgrounds in order to delete the "theme" information contained in the backgrounds. Distortions were created with the following criteria in mind: nonsense backgrounds should not look like the original backgrounds; subjects should not be able to name these backgrounds; roughly the same number of lines and angles should be employed in the nonsense background as in the original; nonsense backgrounds should preserve a three dimensional quality. The nonsense backgrounds resulting from this set of criteria did not appear as a random set of line segments, but as a "coherent" background that lacked any theme information. An informal pilot test of the backgrounds indicated that subjects could not attach a name to the backgrounds. Even pilot subjects who were familiar with the original backgrounds were unable to identify which nonsense background was constructed from which original background. It was important to preserve the three dimensionality of the nonsense backgrounds in order to provide planes on which objects could be supported. This was achieved with all 16 nonsense backgrounds. The objects were situated in the nonsense backgrounds in exactly the same places as they were in the no background conditions of the previous studies.

Design

The design of Experiment 3 was essentially the same as Experiment 1. The only difference was that instead of no

background control conditions, the nonsense background controls were employed. Each object was cued only once for each subject and the condition in which any given object was cued was counterbalanced across subjects. The correct answer, yes or no, was also counterbalanced. Trials were presented in a random order, with the constraint that no backgrounds were repeated on consecutive trials, and no object sets were repeated on consecutive trials.

Apparatus and Procedure

The apparatus and procedure employed in this experiment was the same as Experiments 1 and 2.

Results

As in Experiment 1, there was a main effect of block in all the analyses (i.e. the second time through the scenes subjects performed with fewer errors). However, the block factor did not interact with any of the factors of interest so in the analyses to be reported here, the data has been collapsed across block and block has been dropped as a factor.

Accuracy

As with Experiments 1 and 2 accuracy was assessed by percent correct and A'. In order to assess the role of

consistent and inconsistent backgrounds a two factor within subjects Analysis of Variance was performed on the data. The overall percent correct was 62.0%. Subjects were 3.6% more accurate in the Background Consistent condition than the matching Nonsense Background condition and were 1.5% less accurate in the Background Inconsistent condition than the matching Nonsense Background. This Consistency x Background Type interaction was significant, $F(1,23) = 4.33$, $p = .0487$. Although the pattern of data was similar for the A' measure, the Consistency x Background Type interaction was not significant, $F(1,23) = 2.04$, $p = .1671$.

Table 7

Mean Percent Correct and Mean A' on Object Identification Task for Experiment 3.

	Background Consistent	Background Inconsistent
Meaningful Background	65.6% (.721)	59.4% (.675)
Nonsense Background	62.0% (.670)	60.9% (.676)

Note - A primes in parentheses.

The primary purpose of conducting Experiment 3 was to

determine if the good performance in the Background Consistent condition was due to facilitation or whether the poor performance in the Background Inconsistent condition was due to interference. In order to assess this, simple t-tests were performed between the Background Consistent condition and its matching Nonsense Background condition, and between the Background Inconsistent condition and its Nonsense Background control. As can be seen in Table 7, the difference between the Background Consistent condition and the Nonsense Background condition was 3.6%. This difference was significant, $t(23) = 2.096$, $p < .05$. The difference between the Nonsense Background and the Background Inconsistent condition was 1.5% and this was not significant. The difference between the Background Consistent and the Nonsense Background on the A′ measure was .061. This difference was significant with a one tailed t-test, $t(23) = 1.953$, $p < .05$. The difference between the Nonsense Background and the Inconsistent Background condition was .001 for the A′ measure. This difference was not significant.

Error Data - Misses and False Alarms

The percentage of misses and false alarms for each condition can be seen in Table 8. Overall, subjects missed on 48.8% of the yes trials and false alarmed on 27.2% of the no trials. ANOVAS were performed on both the miss and false

alarm data. The miss results will be discussed first.

Table 8

Mean Percent Misses and False Alarms for Experiment 3.

	Background Consistent	Background Inconsistent
Meaningful Background	38.6% (30.8%)	57.3% (22.9%)
Nonsense Background	47.4% (28.7%)	51.8% (26.6%)

Note – False Alarms in parentheses.

The two significant effects in the miss data were a main effect of consistency $F(1,23) = 17.84$, $p = .0003$, and the Consistency x Background type interaction, $F(1,23) = 19.52$, $p = .0002$. As can be seen in Table 8 subjects missed more frequently in the background inconsistent and the nonsense background conditions than in the background consistent conditions.

Analysis of the False Alarm data revealed a main effect of Consistency, $F(1,23) = 4.97$, $p = .0359$ as the only significant effect.

Discussion

The results of the Consistent and Inconsistent Background conditions in Experiment 3 replicated the findings of Experiments 1 and 2. The most interesting result from the current experiment, however, is the fact that backgrounds that are episodically consistent seem to be facilitating object identification while backgrounds that are episodically inconsistent appear not to interfere.

It thus appears that the no background conditions employed in Experiments 1 and 2 were not the appropriate baseline upon which to increment facilitation versus interference. The good performance on no background conditions was apperantly due to the fact that these scenes were perceptually easier. The nonsense background employed in the current experiment is a more reasonable control condition. Subjects still had to identify where the objects were located in the scene and parse the objects from their background, but attention and capacity spent processing the background would not have aided the subject in the object identification task.

One assumption about the processing of scenes must be made, however, if one is to believe that the nonsense backgrounds were an adequate control. The assumption is that processing of the background must go on in parallel, to some extent, with the processing of the objects. If a strictly serial model of processing is assumed, (one that

says identify backgrounds first, don't continue until background has been identified, and then identify objects with remaining time) then it could be argued that all of the time when the scene was visible was spent attempting to make sense out of the nonsense background, and thus the subject never had time to identify the objects. However, if this were the case subjects should not have been any better than chance in the nonsense background condition and they did perform significantly better than chance.

The pattern of "benefit" from consistent backgrounds and no "cost" from the inconsistent backgrounds is interesting from an ecological validity viewpoint. Had the results been in the opposite direction, that is, all cost for inconsistent background and no benefit for consistent background, then one might have argued that the results were less interesting. After all, how often in "real-world" perception do we see refrigerator objects in the bedroom, or vice versa? The results obtained from Experiment 3 fit nicely with how one might expect the perceptual system to work; that is, consistent information facilitates object identification, but inconsistent information does not interfere. If inconsistent information interfered then one would be at a real disadvantage in identifying an object when that object was inconsistent. This would predict that if a lion were in your living room it would take you a long time to identify that it was a lion. This organization

would not be very adaptive.

It is difficult to acertain whether benefit from consistent information with no cost for inconsistent information is predicted by schema activation, because it has not been made explicit in models of schema activation whether activating the wrong schema interferes or whether activating the right schema facilitates, or both.

The pattern of benefit from consistent information and no cost from inconsistent information was also observed in fixation duration data reported by Henderson et al. Henderson et al. found that having a related object in the fovea prior to fixating the target object resulted in faster naming times to the target object. Furthermore, he found that having an unrelated object as the prime did not interfere with naming as compared to a neutral baseline. This issue of facilitation versus inhibition will be discussed further in the general discussion.

CHAPTER 5

GENERAL DISCUSSION

All three experiments reported here demonstrated that
scene backgrounds play a role in object identification when
the scene is presented for a brief duration: the degree to
which a background is consistent with the target object
predicts performance on the object identification task. In
addition, Experiment 2 indicated that there is no effect of
the surrounding, non-cued cohort objects on this process.
Finally, Experiment 3 showed that the relatively good
performance in the Background Consistent condition is the
result of facilitation and the relatively poor performance
in the Background Inconsistent condition is not the result
of interference.

These results (particularly the results of Experiment
2) pose some problems for the object relation route to
schema activation as proposed by Biederman et al. (1982).
Experiment 2 did not replicate his finding that subjects
perform worse on the object identification task if the
target object is undergoing a violation. I would conclude
from this that Biederman's conception of what activates a
schema and how that schema operates is not quite accurate.

In the introduction, I outlined Biederman's model for
schema activation. He assumes that some relations between
objects can be understood before the objects are fully

identified (i.e. the syntactic relations, support and interposition). I have argued that this is not a valid assumption. Biederman et al. claim that these relations between objects are responsible for activating the appropriate schema for the scene. Furthermore, they claim that the schema that is activated then facilitates further identification of objects. They assert that when an object is undergoing violation, the correct schema can be accessed but that this schema activation interferes with identifying the object undergoing violation. Presumably this interference occurs because the schema dictates not only what objects can occur in the scene, but where those objects are likely to occur. Experiment 2 in the present series of studies does not provide any evidence that object relation information dictated which schema should be activated and that this schema interfered with the identification of the cued object. If this were the case, subjects should have performed worse in the background inconsistent condition when the target object was not with its cohorts than when it was with its cohorts. This was not the case. However, it should be noted that Experiment 2 was only able to address the relation of probability. The other relations, support, interposition, position, and size were not directly manipulated in any of the present experiments.

The question then presents itself, can we modify Biederman's schema activation model to account for the data

presented here? The obvious possibility is that the background activates the schema instead of object relations. Perhaps subjects rapidly identify the background as well as partially identify some of the objects. This information, mostly background information, accesses a schema for this scene and the schema facilitates object identification. The evidence presented here may be consistent with this view, but does not exclusively point to this interpretation.

The backgrounds employed in these three experiments were fairly constrained in that the theme of the scene was clear from the background. By this I mean that the kitchen sink scene was constructed in such a way that it could only be a kitchen sink and would not be confused with a bathroom sink or with any other background. But what would happen if the backgrounds were less constrained so that the information obtained from identifying the background did not necessarily predict what the scene was about? If schema activation is a result of correctly identifying the background, then one of three things could happen if scenes were constructed with "generic" backgrounds. When presented with nondescript backgrounds subjects might settle on one interpretation of the background, access that schema and if they were right, facilitation should occur, whereas if they are wrong the schema should interfere. This option does not seem very plausible because they would on average be wrong more than right. Another possibility is that the generic

background may access schemas for all scenes that could plausibly have that background. Perhaps this reduces the activation of any one schema so the effects of schema facilitation or interference should be less than with descriptive backgrounds. Finally, schemas may only be activated when the backgrounds are complete enough to fairly accurately access the correct schema. In this case there should be no effect of generic background over the nonsense background. Unfortunately, the data from the current experiments cannot address this issue so a definitive answer must wait until the appropriate experiment is done.

To summarize, although we can not rule out that schema activation might have been playing a role in the current experiments, we can conclude that Biederman's route to schema activation may not be correct. At this point I would like to move on and discuss the ramifications of these data for the entirely bottom-up approach to scene context effects as outlined by Henderson et al.

Henderson et al. has provided evidence that an object fixated prior to fixating a target object can facilitate identification of the target object if the previously fixated object was semantically related (i.e. dog primes cat, coat primes hat). The evidence from the current experiments does not entirely support his claim that these sorts of priming effects can account for scene context effects. There are some reasons however, why the current

results may differ from those presented by Henderson et al. In the following section I will explain these differences and the extent to which these differences might or might not affect the conclusions drawn by Henderson.

In its purest form, the object to object priming hypothesis would predict that there should be no effect of background on the object identification task, since the only facilitation should result from other non-cued objects priming the target object. The data presented in these three experiments is in direct conflict with this hypothesis. The major finding was the influence of background information on object identification, and there was no effect of the cohort or non-cued objects in Experiment 2. As I indicated in the discussion of Experiment 1, we can reconcile the first finding (role of backgrounds) with the Henderson model if we assume that backgrounds are operating much as objects do, only that they are easier to identify because they are so large. However, this modification does little to help explain why we obtained no priming from the cohort set in Experiment 2. For this reason I would like to further explore some of the differences in methodology to try account for these apparently conflicting results.

One possibility is that the measures Henderson employed (fixation duration, naming time) and the measure used in the current experiments (yes/no response to object

identification question) tap different stages in processing. Henderson et al. (1987) acknowledge this possibility and claim that the object identification task employed in Biederman's research and the present research may tap processing at a scene integration stage rather than at an object identification stage. That is, objects that are consistent with the scene context are more easily integrated into the representation of the scene and therefore less likely to be lost from a memory buffer in which the presence or absence of the target object is read from. It is not possible in the context of the current experiments to localize the effects to a specific point in the time course of processing.

Henderson et al. also claim that, to the extent to which the priming mechanism is automatic, it should occur in all scene processing even if post-perceptual processing is occurring. Evidence for this automatic priming between related objects was not observed in the current research. However, the objects may have to be fixated (or at least be close to fixation) in order for the priming to occur. In the current experiments none of the cohort objects were fixated and all were greater than 2 degrees from fixation.

On the other hand, the evidence obtained from Experiment 3 does suggest that automatic processing is occurring. To the extent there was a benefit and no cost related to the episodic appropriateness of the background

indicates that these effects are not due to some conscious problem solving strategy. If a conscious problem solving strategy was employed by the subject that resulted in the subject making predictions about what was present in the scene then there should have been interference effects in the Background Inconsistent condition as well as facilitation in the Background consistent condition.

An assumption that must be made in order for object to object priming to account for scene context effects is the assumption that objects which fit in the scene are more likely to be semantically related to one another than objects that could not fit in the scene. I am not convinced that this assumption is completely valid. The present experiments did not demonstrate any cohort effects. For the sake of argument, however, lets assume that a cohort effect would be observed if the objects had been fixated. For example, a doll and a suitcase both appeared in the bedroom scene of this experiment and were responded to as if they were consistent with the scene context (i.e. subjects made relatively few errors). A carton of milk was put in the bedroom background in the Background inconsistent condition and subjects responded to the carton of milk as though it was inappropriate in this context (subjects made more errors). It is hard for me to believe that "suitcase" and "doll" are more semantically related than "carton of milk" and "doll". So it seems as though the validity of the

priming mechanism for picture perception is dependent upon how one defines what should prime what. Many objects used in the current experiments may not be considered "semantically related", meaning that they are not from the same semantic category. Furthermore, most objects in these experiments are not strong associates of one another (as were the objects used by Henderson). However, Henderson used semantic associations between objects and this may not be appropriate. Perhaps the underlying principle concerning what should prime what is the frequency of co-occurence of the objects in the real-world.

In conclusion, there are some reasons why Henderson et al. would have observed priming effects that were not apparent in the current data. However, I think the current data argue that Henderson's object to object priming mechanism cannot account for all of the scene context effects. First, it seems unlikely that background information plays such an important role at, perhaps, a later integration stage, without it having some effect at an earlier processing stage. Second, the priming mechanism is based upon the assumption that objects co-occurring in a scene are semantic primes of one another. Thus, if we were to present scenes in which the objects were not semantic primes of one another but could co-occur in the scene, we should not get context effects. Evidence from less on-line measures seems to indicate that context effects can occur

even when the objects are not semantic primes of one another. This leaves us with two issues: can a priming mechanism account for the context effects if we somehow re-define what should prime what; and how do we determine what should prime what? The data presented here do little to answer these questions other than to suggest that a model based on priming would have to account for backgrounds priming objects and the priming network should be organized according to episodic co-occurence, rather than semantic relatedness.

Up to this point I have tried to indicate how two extremely different models of scene perception would have to be altered in order to account for data from the current experiments. The route to schema activation would have to be changed from object relations to some other route. The current data suggest that, at least when the backgrounds are constraining enough and when there are no foveal objects, then background identification may be the route to schema activation. Furthermore, the data suggest that once the appropriate schema is activated, the effect of the schema is only facilitory and not inhibitory. In order for a priming mechanism to account for the current data, the priming network would have to be organized on the principle of episodic co-occurence, not semantic relatedness. Furthermore, it must be assumed that backgrounds can operate like large objects in this episodically arranged network.

The current data cannot distinguish between the revised schema model or the revised priming model. This must await further research.

The research presented here has contributed to our knowledge about scene processing. First, we have demonstrated that "schema-like" effects can be achieved even when there are no diagnostic objects or diagnostic features in the fovea. The background effects obtained in all three of the experiments reported here indicate that subjects can identify backgrounds even though in order to do so they must process information that is presented peripherally as well as foveally. Second, it appears that context effects due to background are facilitory in nature and not inhibitory. Future models of scene processing should not only account for context effects in terms of other objects in the scene, but should explicitly include the facilitating role that backgrounds can play in object identification.

Scene name in parentheses Indicates scene that was used for background inconsistent conditions.

* Indicates non-cued object

Scenes and Objects

Distractors

1. Bedroom
   (Refrigorator)

   1. Teddy Bear
   2. Doll
   3. Suitcase   1. Radio
   4. Baseball Cap   2. Alarm Clock
   5. Pennent *

2. Broom Closet
   (Desk)

   1. Iron
   2. Scrub Brush
   3. Paper Towels   1. Vaccuum Cleaner
   4. Bucket   2. Dust Pan
   5. Wisk Broom *

3. Clothes Closet
   (Oven)

   1. Pants
   2. Bowtie
   3. Glove   1. Coat
   4. Shoe   2. Sock
   5. Hat *

4. Construction
   (Porch)

   1. Drill
   2. Hammer
   3. Saw   1. Screwdriver
   4. Saw Horse   2. Wrench
   5. Ladder *

5. Desk
   (Broom Closet)

    1. Briefcase
    2. Lamp
    3. Phone   1. Pad of Paper
    4. Stapler   2. Ruler
    5. Picture *

6. Diner
   (Shower)

    1. Coffee Cup
    2. Fork
    3. Salt Shaker   1. Menu
    4. Syrup Pitcher   2. Plate
    5. Ketchup *

7. Fireplace
   (Laundry)

    1. Candle
    2. Chair
    3. Clock   1. Matches
    4. Logs   2. Poker
    5. Bellows *

8. Refrigerator
   (Bedroom)

    1. Butter
    2. Cheese
    3. Lettuce   1. Six Pack
    4. Milk   2. Egg Carton
    5. Onion *

9. Laundry
   (Fireplace)

    1. Laundry Basket
    2. Laundry Soap Box
    3. Hanger   1. Clothes Pin
    4. Shirt   2. Hamper
    5. Bleach *

10. Oven
    (Clothes Closet)

    1. Sauce Pan
    2. Spoon
    3. Teapot   1. Skillet
    4. Turkey   2. Spatula
    5. Oven Mit *

11. Pool
    (Street)

    1. Flipper
    2. Raft
    3. Grill   1. Beach Chair
    4. Beach Ball   2. Snorkel
    5. Life Saver *

12. Porch
    (Construction)

    1. Birdhouse
    2. Rollerskate
    3. Pumpkin   1. Cat
    4. Newspaper   2. Flower Pot
    5. Watering Can *

13. Kitchen Sink
    (Toilet)

    1. Coffee Maker
    2. Eggbeater
    3. Toaster   1. Sponge
    4. Wine Glass   2. Dish Rack
    5. Dish Detergent *

14. Shower
    (Diner)

    1. Bath Mat
    2. Slippers
    3. Soap   1. Shower Cap
    4. Towel   2. Wash Cloth
    5. Shampoo *

15. Street
    (Pool)

    1. Bike
    2. Fire Hydrant
    3. Mailbox   1. Stop Sign
    4. Parking Meter   2. Traffic Light
    5. Wagon *

16. Toilet
    (Kitchen Sink)

    1. Toilet Brush
    2. Toilet Paper
    3. Kleenex   1. Air Freshener
    4. Toilet PLunger   2. Waste Basket
    5. Baby Powder *

REFERENCES

Antes, J. R. (1977). Recognizing and localizing features in brief picture presentations. Memory and Cognition, 5, 155-161.

Antes, J. R. & Penland, J. G. (1981). Picture context effects on eye movement patterns. Fisher, Monty & Senders Eye movements: Cognition and Visual Perception. Hillsdale: Earlbaum, 1981.

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. Psychological Research, 43, 277-292.

Biederman, I. (1972). Perceiving real-world scenes. Science, 177, 77-80.

Biederman, I. (1981). Do background depth gradients facilitate object identification? Perception, 10, 573-578.

Biederman, I. (1982) Final Report to U.S. Army Research Institute.

Biederman, I., Glass, A., & Stacy, E. (1973). Searching for objects in real world scenes. Journal of Experimental Psychology, 1, 22-27.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. Cognitive Psychology, 14, 143-177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. Journal of Experimental Psychology, 103, 597-600.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology, 3, 316-355.

Grier, J.B. (1971). Nonparametric Indexes for sensitivity and bias: computing formulas. <u>Psychological Bulletin</u>, <u>6</u>, 424-429.

Henderson, J. M., Pollatsek, A., & Rayner K. (1987). The effects of foveal priming and extrafoveal preview on object Identification. <u>Journal of Experimental Psychology: Human Perception and Performance</u>, <u>3</u>, 449-463.

Loftus, G. R. & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. <u>Journal of Experimental Psychology: Human Perception and Performance</u>, <u>4</u>, 565-572.

Mackworth, N. H. & Morandl, A. J. (1967). The gaze selects Informative details within pictures. <u>Perception and Psychophysics</u>, <u>2</u>, 547-552.

Marr, D. (1982). <u>Vision : A computational Investigation Into the human representation and processing of visual Information</u>. San Francisco : W. H. Freeman Company.

Nelson, W. W. & Loftus, G. R. (1980). Functional visual field during picture viewing. <u>Journal of Experimental Psychology: Human Perception and Performance</u>, <u>6</u>, 391-399.

Parker, R. E. (1978). Picture processing during recognition. <u>Journal of Experimental Psychology: Human Perception and Performance</u>, <u>4</u>, 284-293.

Salda, S. & Ikeda, M. (1979). Useful visual field size for pattern perception. <u>Perception and Psychophysics</u>, <u>25</u>, 119-125.